

2004年度 修士論文

可変長セグメントパターンマッチング に基づく楽音の音高・楽器推定

Musical Instrument Identification based on
Pattern Matching of Variable Length Segment

提出日 2005年2月2日

指導教授：白井 克彦 教授

早稲田大学大学院 理工学研究科 情報・ネットワーク専攻

学籍番号：3603U093-2

椿 雅也

TSUBAKI Masaya

目 次

| | | |
|-------|-----------------------------------|----|
| 第 1 章 | 序論 | 1 |
| 1.1 | 研究の背景 | 1 |
| 1.2 | 従来研究 [5] | 2 |
| 1.2.1 | 単音抽出および音高同定 | 2 |
| 1.2.2 | 音源同定 | 3 |
| 1.2.3 | 音源弁別 | 4 |
| 1.2.4 | リズム同定 | 4 |
| 1.3 | 研究概要 | 6 |
| 第 2 章 | 楽器・音高識別処理 | 8 |
| 2.1 | 処理概要 | 8 |
| 2.2 | 楽器・音高識別処理方法 | 9 |
| 2.3 | 時間セグメントにおける楽器・音高識別処理方法 | 10 |
| 2.3.1 | 手法 1:状態スペクトルを用いた時間セグメントゲイン値可変手法 | 11 |
| 2.3.2 | 手法 2:状態スペクトルを用いた時間セグメントゲイン値一定手法 | 13 |
| 2.3.3 | 手法 3:スペクトル時系列を用いた時間セグメント内ゲイン値一定手法 | 15 |
| 第 3 章 | 最適時間セグメント抽出処理 | 16 |
| 3.1 | 処理概要 | 16 |
| 3.2 | 抽出のための計算方法 | 17 |
| 第 4 章 | 識別実験 | 19 |
| 4.1 | 実験概要 | 19 |
| 4.2 | オーバーラップ | 19 |
| 4.3 | テストデータ | 20 |

| | | |
|------------|----------------------------|-----------|
| 4.4 | 楽器データ | 21 |
| 4.4.1 | 単音データ（混合音モデル作成用） | 21 |
| 4.4.2 | 楽音データ | 21 |
| 4.4.3 | 編成バリエーション楽曲 | 21 |
| 4.5 | 実験条件 | 22 |
| 第5章 | 実験結果 | 24 |
| 5.1 | 実験 1-1 結果 | 24 |
| 5.2 | 実験 1-2 結果 | 26 |
| 5.3 | 実験 2-1 結果 | 28 |
| 5.4 | 実験 2-2 結果 | 30 |
| 5.5 | 平均スペクトル距離 | 32 |
| 第6章 | 考察 | 33 |
| 第7章 | まとめ | 35 |
| | 参考文献 | 36 |

目 次

| | | |
|------|----------------------------------|----|
| 1.1 | 処理の全体像 | 6 |
| 1.2 | 時間セグメント | 7 |
| 1.3 | 楽器・音高識別処理概念図 | 7 |
| 2.1 | 楽器・音高識別イメージ | 9 |
| 2.2 | 状態区間抽出 | 11 |
| 2.3 | 楽器・音高識別処理での DP マッチング | 12 |
| 2.4 | 状態スペクトルを用いた時間セグメントゲイン値一定手法のイメージ図 | 14 |
| 3.1 | 最適時間セグメント抽出処理 | 16 |
| 3.2 | 区間距離計算 | 17 |
| 4.1 | オーバーラップの例 | 20 |
| 4.2 | 実験 1-1、1-2 用テストデータ | 23 |
| 4.3 | 実験 2-1、2-2 用テストデータ | 23 |
| 5.1 | 実験 1-1 手法 1 | 24 |
| 5.2 | 実験 1-1 手法 2 | 25 |
| 5.3 | 実験 1-1 手法 3 | 25 |
| 5.4 | 実験 1-2 手法 1 | 26 |
| 5.5 | 実験 1-2 手法 2 | 27 |
| 5.6 | 実験 1-2 手法 3 | 27 |
| 5.7 | 実験 2-1 手法 1 | 28 |
| 5.8 | 実験 2-1 手法 2 | 29 |
| 5.9 | 実験 2-1 手法 3 | 29 |
| 5.10 | 実験 2-2 手法 1 | 30 |

| | |
|----------------------------|----|
| 5.11 実験 2-2 手法 2 | 31 |
| 5.12 実験 2-2 手法 3 | 31 |

表 目 次

| | | |
|-----|-----------------------------------|----|
| 1.1 | 自動採譜に関連する主な研究例における処理の比較 | 5 |
| 5.1 | 平均スペクトル距離 | 32 |

第1章 序論

1.1 研究の背景

近年、電子楽器やコンピュータの普及、進歩によって容易に音楽データを扱えるようになり、その結果コンピュータを利用したさまざまな音楽活動が行われている。我々の日常生活においても、数多くの電子音が存在し、新たな電子音の登場には同時に質の向上も求められている。カラオケの音や携帯電話の普及による着メロの音がその一例である。またDTMという言葉も知られているように、コンピュータによる演奏や作曲や編曲などが一般的になった。それらの活動の支援として、コンピュータで音楽から楽音を認識し楽譜を作成する自動採譜システムの研究も行われている。

現在、携帯電話の着メロやカラオケなどの音データを作る場合、人間が原曲を聴いて音高などの情報をMIDIデータとして記録する、いわゆる耳コピによる手法が主流である。MIDIデータを作るために、作業者は楽音の各パート毎の楽譜を作っている。そして、各パート毎の楽譜を作るためには、楽器を識別し、楽音の旋律を和音毎の時間に分割し、その時間分割区間で鳴っている和音の各音高とゲイン値を決定しなければならない。

本研究ではその過程に着目し、楽音を入力として、楽音の音響信号から楽器名、音高、ゲイン値を決定し、MIDIデータを出力することを目的とする。これらが達成されれば、楽音からのMIDIデータの自動作成のみならず、普段音楽を聴く場合に特定の楽器音のボリュームを調整したり、音色を変更することで、個人の趣向に合った音楽鑑賞ができるようになる。

そのために各楽器、各音高の単音から混合音のテンプレートを作成し、入力音響信号とのマッチングを行う。その際、混合音テンプレートの各単音のパラメータを同時に決定し、その値をその単音の強さ(ゲイン値)とし、入力音響信号に含まれる楽器名、音高、ゲイン値を決定する。また、楽音の旋律を和音毎の時間に分割するために、マッチング距離を用いて楽音全体で最小累積距離となるように分割区間を求める。

1.2 従来研究 [5]

楽音の音響信号から、その入力楽曲の MIDI データを出力する処理として、自動採譜処理というのが考えられる。自動採譜処理の具体的な内容は、音符列、ビート、和音 (コード) などの記号表現的な事象を抽出する処理がまず考えられる。このような抽出を行い、さらに結果を楽譜の形に表す処理は採譜 (transcription) と呼ばれる。他にも、情感や演奏技法といった、譜面に現れない抽象度の高いものを認識対象として考えることも可能である。本研究では自動採譜処理における音符列、ビート、和音 (コード) などの記号表現的な事象を抽出する処理を利用して MIDI データの自動生成を目指す。

自動採譜処理における音符列、ビート、和音 (コード) などの記号表現的な事象を抽出する処理を実現するにあたり、必要となる機能がいくつか存在する。ここではそれらについて従来研究で用いられている手法を概観する。

表 1.1 は、それぞれの手法が各要素技術をどのように実現しているかを示した。

1.2.1 単音抽出および音高同定

入力音響信号から、そこに存在する単音を抽出する処理である。楽譜で表した際の音符に相当する。ここでは音符の生起時刻の抽出のみを考える。この処理には、以下のような手法が提案されている。

周波数解析による手法

FFT や、フィルタバンクを用いて周波数解析を行う処理である。FFT を用いる方法は、自動採譜に関する研究が行われた当初から利用されている。また、FFT では周波数軸上のサンプル点が線形に並ぶため、音楽音響信号の解析には不向きである。そこで、フィルタをその中心周波数が対数軸上に並べたフィルタバンクを用いる方法がある [17]。

単音を抽出するには、これらの周波数解析の結果からそのパワーのピークを時間方向に追跡することで周波数成分を抽出し、それらを単音に相当するクラスターへとクラスタリングする必要がある。ここで、クラスタリング処理においては、楽器音の多くが長波構造を持つことを利用する例が多い。

また、片寄らは、あらかじめ音源ごとに高調波のパワー比のテーブルを用意し、単音の抽出に応用している。この手法では、単音の抽出と同時にその周波数が検出でき、音高の同定が可能になる [15]。

音響信号のパワー値を用いる手法

音響信号の振幅を用いて、短時間区間におけるパワー値を計算し、その増加と減少を検出することで単音の立ち上がりと立ち下りを抽出する手法である。この手法は比較的単純な処理で実現が可能であるが、同時に複数音が存在する区間では、ある単音の立ち上がりが他の単音のエネルギーによって隠蔽されることが多く、特に立ち上がりの弱い音源に由来する単音の抽出は困難である。

また、音高の同定を行う場合には、単音が存在する区間を切り出した後に周波数解析を行う例が多い。

楕形フィルタによる手法

楕形フィルタによって、長波構造を保ったまま単音に相当する信号を抽出する。三輪らはこの手法による抽出を試みている [16]。この手法では、周波数解析を行わずに処理が可能であるため、周波数成分のクラスタリングの問題を回避できる。しかしながら、複数の単音が整数倍の周波数比を持っている場合には、楕形フィルタのみではそれらの弁別は本質的に不可能である。

1.2.2 音源同定

抽出された単音が、どの音源に由来するものであるかを同定する処理である。楽譜で表現したときに、音符の属するパートを同定することに相当する。

一般に、音源名を得るためには、音源名とそれに対応したテンプレートを準備し、知識として用意する必要がある。ここに挙げる各手法でも、それぞれにおいてテンプレートを利用している。

聴覚心理学においても、計算機上での処理に適用するほどには人間の音源同定機構に関して知見が得られていない。そのため、これまでに様々な手法が試みられている。

周波数成分の特徴量による手法

OPTIMA[17] では、単音抽出の際に得られた周波数成分から物理的な特徴量を抽出している。また、あらかじめ音源ごとに特徴量のデータベースを準備している。これらと比較することで音源同定を行う。同定は、主成分分析によって各特徴量を音色空間と呼ばれる

空間に変換した後、判別分析を応用した処理によって進められている。

周波数成分の特徴量は、その値の音源ごとの差が比較的容易に類推できるため、音源同定の手がかりとして実装しやすい対象となっている。また、ここで挙げる他の手法と比べると抽象度が高く、計算量等の観点からも使いやすい対象である。

テンプレートマッチングによる処理

Ipanema[18]では、音源同定に際して周波数成分の抽出を行わずに、音源ごとに用意した波形テンプレートと、入力信号から切り出された波形との比較をする。

一般に、周波数成分を用いる場合と比較して、波形レベルでは同一音源による信号であっても、楽器固体や音階、音量などによる差が大きい。そこで、この手法では、FIR フィルタを用いてテンプレートを変形し、また入力信号との位相の差も吸収した上で、最も差の少なかった音源名を出力する。

パワー値の時間変動による手法

三輪らの手法[16]では、入力信号とテンプレート信号について波形レベルで音響パワーの時間的な推移を調べ、それらを比較することで音源名を得る。

1.2.3 音源弁別

2つ以上の単音に対して、同一の音源に所属するものごとにグループ化する処理を指す。音源同定と並んで必要な処理である。音源同定が行われればこの処理は完了するが、実際には音源同定が困難である場合が多く、そのような場合には音源同定に先んじて音源弁別を行うことで、音源同定処理を容易にすることが可能となる。これまでの研究では、音源同定を行わずに音源弁別を行った例は見られない。

1.2.4 リズム同定

入力の音楽演奏に含まれているリズムを同定し、拍位置を抽出する処理である。

単音位置情報に基づく手法

既に抽出されている単音、あるいは単音に相当するエネルギーの立ち上がり位置を元に拍位置の同定を行う手法である。音響信号から周波数成分を抽出した上でその立ち上がり時刻を用いてリズムを抽出する例として OPTIMA[17] がある。また、MIDI 信号に含まれる単音の立ち上がり時刻を用いる例として、Machine Rhythm[7] が挙げられる。

単音位置を用いない手法

単音に相当する対象を抽出せずに、リズムを抽出する。片寄は音響パワーの立ち上がりの位置を用いている [15]。また、ドラムの位置やコードの変化を用いてリズムを抽出する手法も提案されている [8]。

いずれの手法でも、リズムそのものの抽出の他にその結果を単音抽出へ応用することが可能である [17]。

表 1.1: 自動採譜に関連する主な研究例における処理の比較

| 手法 | 単音抽出 | 音源同定 | 時間処理 |
|------------|---------------|-------------|-------------------------------|
| 片寄 | パワー分布のみ | | リズム抽出による次音位置推定 |
| OPTIMA | フィルタバンク | 周波数成分特徴量 | リズム抽出による周波数成分整形・コード進行に基づく仮説修正 |
| MRF-OPTIMA | 同上 | 同上 | 上記 + 単音遷移 |
| Ipanema | (周波数解析 (未実装)) | テンプレートマッチング | 単音連繋確率ネットワーク |
| 三輪 | 楕形フィルタ | 信号波形包絡線 | |

1.3 研究概要

本研究では楽音を入力として、従来研究と違った手法を用い、その楽音に含まれる演奏情報を抽出し、入力した楽音に近いMIDIデータを作成することを目的としている。本研究で用いる手法では、単音抽出と音源同定を“楽器・音高識別処理”にて同時に行い、時間処理を“最適時間セグメント抽出処理”にて行うことで、入力音響信号から演奏情報を抽出する。従来研究では単音抽出、音源同定を別々に本研究では単音抽出と音源同定を同時に行うことで、

本研究では入力として、複数楽器で演奏され、各楽器が和音を持つ混合音楽音を用いる。そして、その楽音に対し“最適時間セグメント抽出処理”、“楽器・音高識別処理”の2つの連動した処理を施すことにより、入力楽音に含まれる演奏情報（楽器、音高、ゲイン値）を抽出する。そして、この演奏情報に対し“MIDI生成処理”を施すことにより入力楽音からMIDIデータを自動生成させる。本研究で行う処理の全体像を図1.1に示す。

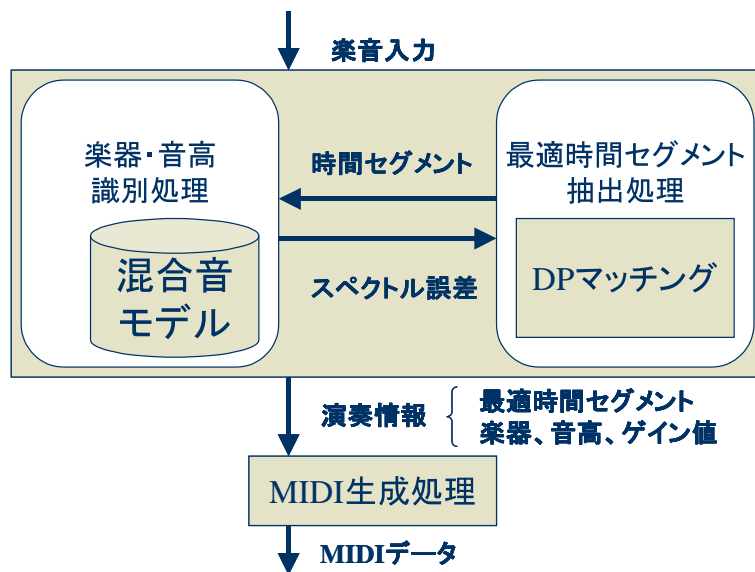


図 1.1: 処理の全体像

“最適時間セグメント抽出処理”では図1.2のような、同じ和音が続く区間（以下、時間セグメント）を抽出する。このように時間セグメントを抽出することで、“楽器・音高識別処理”では時間セグメントでは同じ和音とその時間セグメントの時間長含まれているという条件で楽器・音高識別処理を行っていく。

“楽器・音高識別処理”では“最適時間セグメント抽出処理”で得た時間セグメントにおける音響信号とテンプレート（混合音モデル）とのマッチングを行い、音響信号との距

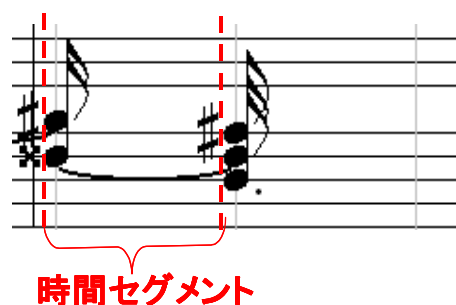


図 1.2: 時間セグメント

離が最小となるようなパラメータを決定する（図 1.3 参照）。

そのときの混合音モデルは複数の楽器の単音モデルから作成され、各単音モデル毎にパラメータが付与されている。これらのパラメータ値を各楽器単音のゲイン値とする。

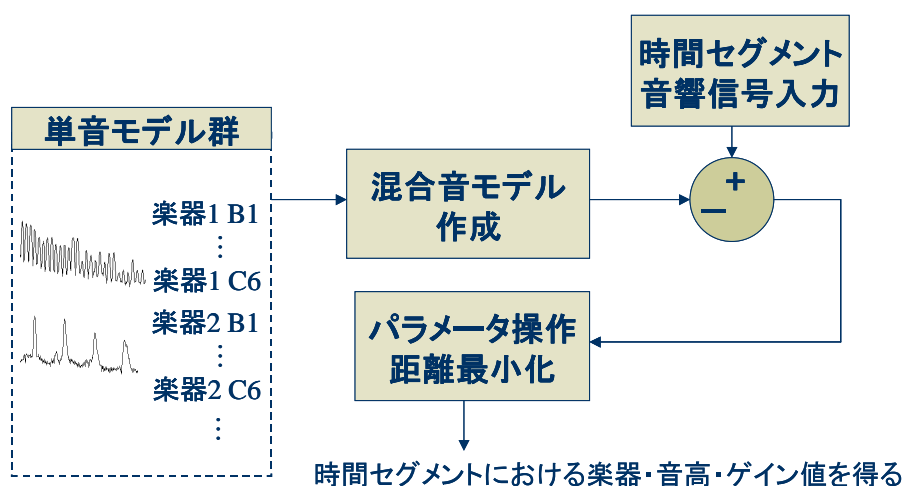


図 1.3: 楽器・音高識別処理概念図

このような“最適時間セグメント抽出処理”、“楽器・音高識別処理”の2つの連動した処理を行うことで、入力楽曲を時間セグメントに分割し、各時間セグメントに含まれる楽器・音高とそのゲイン値を決定することが可能となる。

最後に時間セグメントとして分割した時刻を MIDI データのノート・オン、ノート・オフの時間タイミングと合わせ、得られた楽器、音高、ゲイン値の情報をを用いることで、入力楽音に対する MIDI データを生成する。このように MIDI ファイルを作成する処理を行うのが“MIDI 生成処理”である。

第2章 楽器・音高識別処理

2.1 処理概要

楽器・音高識別処理では 1.3 で述べた同じ和音が続く区間である時間セグメントにおいて処理を施し、時間セグメントに含まれる楽器、音高、ゲイン値を決定する。この時間セグメントは後の第 3 章で述べる最適時間セグメント抽出処理により決定されるものである。この処理の概念図は図 1.3 に示している。

ここでは楽音は単音の集合からなるという考え方から、複数の単音モデルから混合音モデルを構成する。そして、この混合音モデルと時間セグメントにおける音響信号をマッチングさせ、マッチング距離が最小となるように混合音モデルに付与されているパラメータ値を決定する。ここでパラメータは各単音モデル毎に付与されているものであり、これらのパラメータ値を基にして各楽器単音のゲイン値を決定する。また、本研究における楽器・音高識別処理では基本周波数とその倍音構造を用いた音高推定処理を施すのではなく、単音モデルを各音高毎に用意することで楽器と音高の識別を同時に行う。

2.2 楽器・音高識別処理方法

本研究において楽器・音高識別は時間セグメントにおける音響信号と混合音モデルとのマッチング時のパラメータ値の決定によって行われる。図 2.1 が本研究における楽器・音高識別イメージ図である。 s_t は時間セグメントの時刻 t における入力音響信号のパワースペクトル、 c_1, c_2, \dots, c_m は各楽器・音高のパワースペクトル(単音モデル)である。 $g_{t1}, g_{t2}, \dots, g_{tm}$ は各単音モデルに付与されているパラメータ(重み)であり、入力音響信号のパワースペクトルと単音モデルの重み付きの和との距離が最小となるように決定する。この決定されたパラメータ値により入力楽音に含まれる各楽器・単音の成分が分かるのではないかと考えた。

$$s_t \doteq c_1 \times g_{t1} + c_2 \times g_{t2} + \dots + c_m \times g_{tm}$$

図 2.1: 楽器・音高識別イメージ

また、本研究では単音モデルを各音高毎に用意することで楽器と音高の識別を同時に行うので、単音モデルは各楽器の各音高の総数用意する。よって c_m の m はテンプレートに用いる各楽器の各音高の総数である。

ここで、 s_t のパワースペクトルの要素を $s_{t1}, s_{t2}, \dots, s_{tn}$ とおき、単音モデル c_k のパワースペクトルの要素を $c_{k1}, c_{k2}, \dots, c_{kn}$ とおくと、これらは式 (2.1) のように表すことができる。

$$\begin{pmatrix} s_{t1} \\ s_{t2} \\ s_{t3} \\ \vdots \\ s_{tn-1} \\ s_{tn} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{21} & \dots & c_{m1} \\ c_{12} & c_{22} & \dots & c_{m2} \\ c_{13} & c_{23} & \dots & c_{m3} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1n-1} & c_{2n-1} & \dots & c_{mn-1} \\ c_{1n} & c_{2n} & \dots & c_{mn} \end{pmatrix} \cdot \begin{pmatrix} g_{t1} \\ g_{t2} \\ \vdots \\ g_{tm} \end{pmatrix} \quad (2.1)$$

そして、混合音モデルの行列を C とおくと、式 (2.2) の最小 2 乗法による計算式により入力音響信号のパワースペクトルと混合音モデルとの距離が最小となる $g_{t1}, g_{t2}, \dots, g_{tm}$ の

パラメータ列 \mathbf{g}_t を得ることができる。そして、このパラメータ値を基に時刻 t におけるゲイン値を決定する。

$$\mathbf{g}_t = (\mathbf{C}^T \cdot \mathbf{C})^{-1} \cdot \mathbf{C}^T \cdot \mathbf{s}_t \quad (2.2)$$

また、2つのパワースペクトルの距離が最小となる時の混合音モデルとパラメータから生成されるパワースペクトル $\hat{\mathbf{s}}_t$ は以下の式 (2.3) で求めることができ、最小距離 d を式 (2.4) で求めることができる。

$$\hat{\mathbf{s}}_t = \mathbf{C} \cdot \mathbf{g}_t \quad (2.3)$$

$$d = (\mathbf{s}_t - \hat{\mathbf{s}}_t)^2 \quad (2.4)$$

2.3 時間セグメントにおける楽器・音高識別処理方法

2.2 では時間セグメントのある時刻 t における楽器・音高識別処理の方法を述べたが、楽器の音は動的に変化するものであるため、時刻毎に楽器・音高識別を施しても信頼のある結果は望めない。よって、時間セグメントにおけるスペクトル時系列を対象として、楽器の時間変化による識別手法が必要となる。

そこで、本研究ではその手法として以下の3つの識別手法を提案する。

手法1 状態スペクトルを用いた時間セグメントゲイン値可変手法

手法2 状態スペクトルを用いた時間セグメントゲイン値一定手法

手法3 スペクトル時系列を用いた時間セグメント内ゲイン値一定手法

以下の2.3.1、2.3.2、2.3.3で各手法の詳細を述べる。

2.3.1 手法1:状態スペクトルを用いた時間セグメントゲイン値可変手法

昨年度の我々の研究 [2] から楽器の音色は動的な特徴により特徴付けられることが確認できた。そのことは先行研究 [9] で単音データをアタック区間、サステイン区間、減衰区間と3つの状態区間に分け、各状態区間の特徴量を調べることで楽器識別を行っていることから裏付けされる。

このことから、本手法では単音データから図2.2のように Onset、Attack、Sustain、Decay の状態区間を抽出し、混合音モデルで用いる単音データの各状態区間の平均スペクトルから Onset、Attack、Sustain、Decay 区間における混合音モデルを作成した。ここで Onset は音の波形が立ち上がる前の振幅の小さな波形の区間、Attack は波形の振幅が最大値を取る付近、Sustain は波形の振幅の大きさがおよそ一定で継続している区間、Decay は音の波形が減衰している区間と定義した。

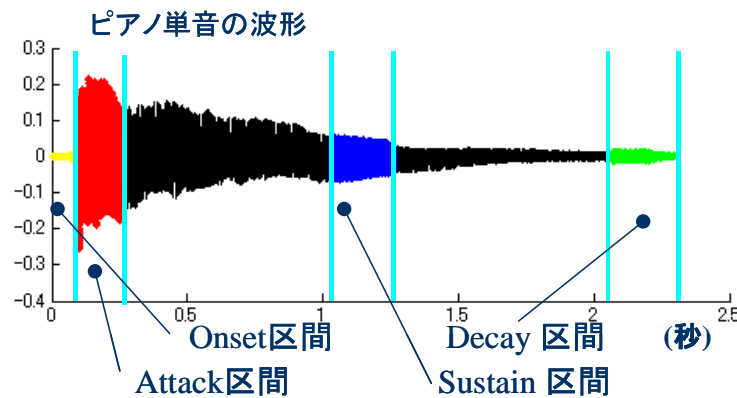


図 2.2: 状態区間抽出

そして、時間セグメントにおける入力音響信号のスペクトル時系列と4状態の混合音モデル(テンプレート)とのDPマッチングを行い、累積距離が最小になるときの経路を求めた(図2.3)。

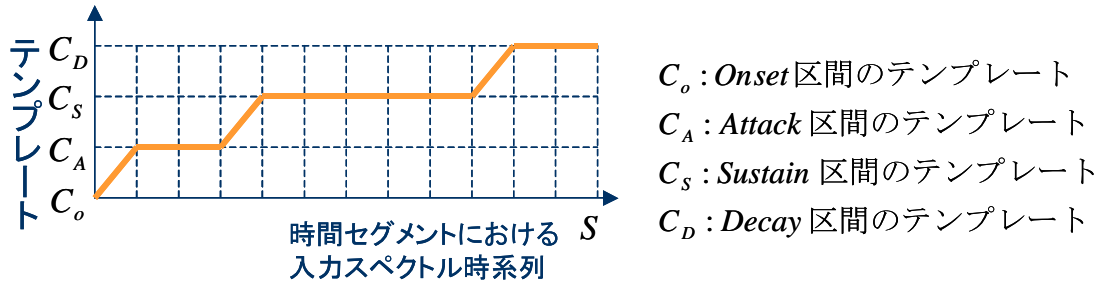


図 2.3: 楽器・音高識別処理での DP マッチング

次に、この経路の各フレームにおけるスペクトルのマッチング時のパラメータ値を取得することで、時間セグメントにおける各単音モデルの各時刻におけるパラメータ値を得ることができる。この場合フレームにおけるスペクトルのマッチングにおいて得られるパラメータ値が各フレーム毎に違った値を取るため、時間セグメントにおいて得られるゲイン値が可変となる。

また、DP マッチングにおける最小累積距離を時間セグメントにおける入力音響信号と混合音モデルとの区間距離 D とする。

2.3.2 手法2:状態スペクトルを用いた時間セグメントゲイン値一定手法

2.3.1 では時間セグメントでのゲイン値が時刻毎に変化してしまう。本来時間セグメントとは一定の和音が続く区間と定義してあるので、時間セグメントにおいての混合音モデルのゲイン値は一定のはずである。よって、ゲイン値を一定にする手法として本手法を提案した。

本手法では時間セグメントにおけるスペクトル時系列との距離が最小となるように、各時間に各状態区間の混合音モデルを配置したものをテンプレートとし、スペクトル時系列に対する最小2乗法で行ことで、時間セグメントにおいて混合音モデルのパラメータ値が一定になるようにした。

式 (2.5) は本手法で用いる計算式である。 $s := \{s_{t_1}, s_{t_2}, \dots, s_{t_l}\}$ は時間セグメントにおける時刻 $t_1 \sim t_l$ におけるパワースペクトル時系列、 C_o 、 C_a 、 C_s 、 C_d はそれぞれ Onset、Attack、Sustain、decay 区間の混合音モデルであり、 $g := \{g_1, g_2, \dots, g_m\}$ は各単音モデルのパラメータである。

$$\begin{pmatrix} s_{t_1} \\ s_{t_2} \\ s_{t_3} \\ s_{t_4} \\ \vdots \\ s_{t_{l-3}} \\ s_{t_{l-2}} \\ s_{t_{l-1}} \\ s_{t_l} \end{pmatrix} \cdot \begin{pmatrix} C_o \\ C_a \\ C_a \\ C_s \\ \vdots \\ C_s \\ C_d \\ C_d \\ C_d \end{pmatrix} \cdot \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{pmatrix} \quad (2.5)$$

図 2.4 は本手法のイメージである。

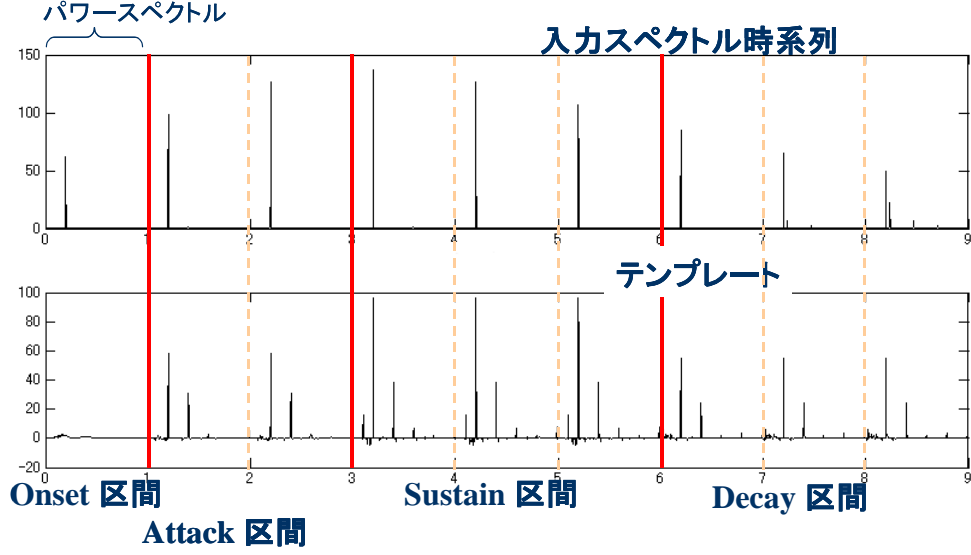


図 2.4: 状態スペクトルを用いた時間セグメントゲイン値一定手法のイメージ図

ここでは最初のフレームを Onset 区間の混合音モデル、2 番目と 3 番目のフレームを Attack 区間の混合音モデル、最後の 3 つのフレームを Decay 区間の混合音モデル、それ以外のフレームを Sustain 区間の混合音モデルとしてテンプレートを $C_2 := \{C_o, C_a, C_a, C_s, \dots, C_s, C_d, C_d, C_d\}$ と定義した。よって、時間セグメントにおけるパラメータ値は以下の式 (2.6) で求めることができる。また、時間セグメントにおけるパワースペクトル時系列の値に近くなるように、テンプレートのパラメータ値を変えたときの値は式 (2.7) で求めることができる。よって時間セグメントにおける区間距離 D は式 (2.8) で求めることができる。

$$\mathbf{g} = (\mathbf{C}_2^T \cdot \mathbf{C}_2)^{-1} \cdot \mathbf{C}_2^T \cdot \mathbf{s} \quad (2.6)$$

$$\hat{\mathbf{s}} = \mathbf{C}_2 \cdot \mathbf{g} \quad (2.7)$$

$$D = (\mathbf{s} - \hat{\mathbf{s}})^2 \quad (2.8)$$

2.3.3 手法3:スペクトル時系列を用いた時間セグメント内ゲイン値一定手法

本手法では Onset、Attack、Sustain、Decay の各区間の平均スペクトルを用いず、単音データのスペクトル時系列をテンプレートとしたマッチングを行う。

式 (2.9) は本手法で用いる計算式である。 $s := \{s_{t_1}, s_{t_2}, \dots, s_{t_l}\}$ は時間セグメントにおける時刻 $t_1 \sim t_l$ におけるパワースペクトル時系列、 $c_{kt_1}, c_{kt_2}, \dots, c_{kt_l}$ は楽器・音高 k の単音データのパワースペクトル時系列である。 $g := \{g_1, g_2, \dots, g_m\}$ は各単音モデルのパラメータである。

$$\begin{pmatrix} s_{t_1} \\ s_{t_2} \\ s_{t_3} \\ \vdots \\ s_{t_{l-1}} \\ s_{t_l} \end{pmatrix} = \begin{pmatrix} c_{1t_1} & c_{2t_1} & \cdots & c_{mt_1} \\ c_{1t_2} & c_{2t_2} & \cdots & c_{mt_2} \\ c_{1t_3} & c_{2t_3} & \cdots & c_{mt_3} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1t_{l-1}} & c_{2t_{l-1}} & \cdots & c_{mt_{l-1}} \\ c_{1t_l} & c_{2t_l} & \cdots & c_{mt_l} \end{pmatrix} \cdot \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{pmatrix} \quad (2.9)$$

ここで、単音モデルとして用意した全楽器・音高のパワースペクトル時系列を用いたテンプレートを C_3 とおく (式 (2.10)) と、時間セグメントにおけるパラメータ値は以下の式 (2.11) で求めることができる。また、時間セグメントにおけるパワースペクトル時系列の値に近くなるように、テンプレートのパラメータ値を変えたときの値は式 (2.12) で求めることができる。よって時間セグメントにおける区間距離 D は式 (2.13) で求めることができる。

$$C_3 = \begin{pmatrix} c_{1t_1} & c_{2t_1} & \cdots & c_{mt_1} \\ c_{1t_2} & c_{2t_2} & \cdots & c_{mt_2} \\ c_{1t_3} & c_{2t_3} & \cdots & c_{mt_3} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1t_{l-1}} & c_{2t_{l-1}} & \cdots & c_{mt_{l-1}} \\ c_{1t_l} & c_{2t_l} & \cdots & c_{mt_l} \end{pmatrix} \quad (2.10)$$

$$g = (C_3^T \cdot C_3)^{-1} \cdot C_3^T \cdot s \quad (2.11)$$

$$\hat{s} = C_3 \cdot g \quad (2.12)$$

$$D = (s - \hat{s})^2 \quad (2.13)$$

第3章 最適時間セグメント抽出処理

3.1 処理概要

最適時間セグメント抽出処理では楽音中から適当な時間を分析区間とし、この分析区間に対して最適な分割を行い、分割された区間を最適な時間セグメント（同じ和音が継続する区間）を決定する（図 3.1）。

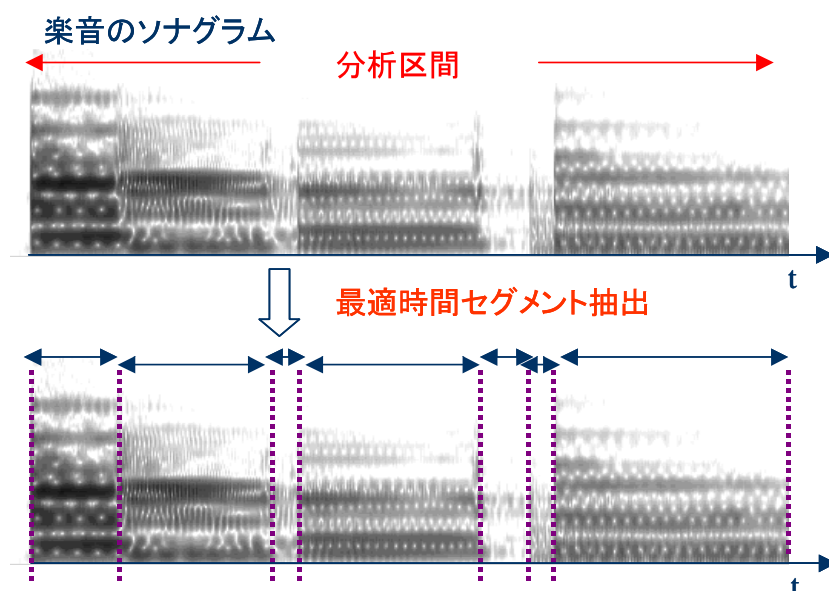


図 3.1: 最適時間セグメント抽出処理

最適時間セグメント抽出処理とは第 1 章で楽器・音高識別処理に同じ和音が継続する区間である時間セグメントを渡す処理であると述べた。しかし、実際の処理では楽器・音高識別処理に渡される時間セグメントは必ずしも同じ和音が継続する区間であるとは限らず、時間セグメントにおいて楽器・音高識別処理を行ったときの区間距離 D を用いることで最適な時間セグメントを決定する。

最適時間セグメント抽出処理では分析区間長を可変長・可変数の時間セグメントで分割

したときに、累積区間距離が最小となるときの分割時刻を求める。そして、その具体的な方法として分析区間における全ての時間セグメントにおいて区間距離 D を楽器・音高識別処理により求め、この D の分析区間での累積の値（累積区間距離）が最小となるような区間の分割を DP マッチングにより行っていく。

このような処理を施すことにより分析区間を同じ和音が継続する区間である最適時間セグメントを抽出でき、さらに時間セグメント毎に含まれる楽器・音高、そのゲイン値を取得することができる。よって、分析区間において MIDI データを作成するための演奏情報を取得することができる。

3.2 抽出のための計算方法

ここでは最適な時間セグメント抽出を行うための計算方法を述べる。

分析区間の時間長を定数 T 、可変長の時間セグメントの最短時間長を定数 T_{min} 、最長時間長を定数 T_{max} とすると、分析区間の最大分割数は $\lfloor \frac{T}{T_{min}} \rfloor$ 、最小分割数は $\lceil \frac{T}{T_{max}} \rceil$ となる。

そして、まず、時刻 0 から時刻 T_{min} までの時間セグメントにおける区間距離 $D(0, T_{min})$ を計算し、次にシフト時間 T_{shift} ずらした時刻 0 から時刻 $T_{min} + T_{shift}$ までの時間セグメントにおける区間距離 $D(0, T_{min} + T_{shift})$ を計算する。このようにシフト時間 T_{shift} ずらしながら時刻 T_{max} まで（図 3.2 の \leftrightarrow の時間）の各時刻における区間距離 $D(0, T_{min}), D(0, T_{min} + T_{shift}), D(0, T_{min} + 2T_{shift}), \dots, D(0, T_{max})$ を計算する。

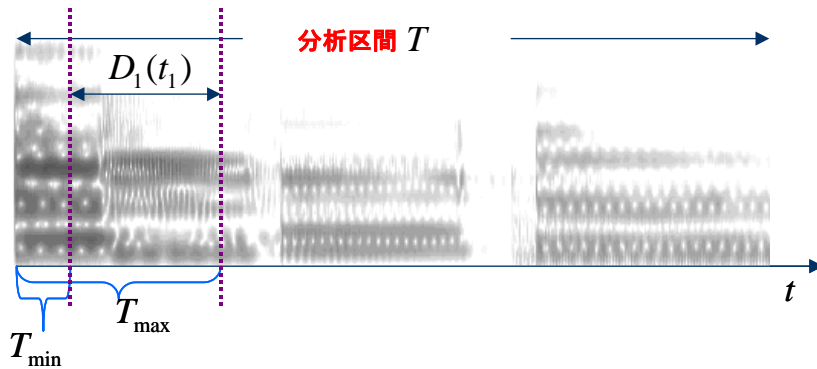


図 3.2: 区間距離計算

次に $D(T_{min}, 2T_{min}), D(T_{min}, 2T_{min} + T_{shift}), \dots, D(T_{min}, T_{min} + T_{max})$ を計算していき、後も同様に $D(2T_{min}, 3T_{min}), D(2T_{min}, 3T_{min} + T_{shift}), \dots, D(2T_{min}, 2T_{min} + T_{max})$ を計算

し、最終的に $D(T - T_{max}, T)$ までの全てにおいての区間距離を求める。

求めた全ての区間距離を用いると時刻 t_n までを n 回分割したときの最小累積区間距離 $D_n(t_n)$ は以下のような漸化式で計算することができる。

ここで t_n の範囲は $1 \leq n \leq \lceil \frac{T}{T_{max}} \rceil$ のとき $nT_{min} \leq t_n \leq nT_{max}$ 、 $\lceil \frac{T}{T_{max}} \rceil < n \leq \lfloor \frac{T}{T_{min}} \rfloor$ のとき $nT_{min} \leq t_n \leq T$ である。

$$\begin{aligned} D_1(t_1) &= D(0, t_1) \\ D_n(t_n) &= \min_{t_{n-1}} [D(t_n - t_{n-1}, t_n) + D_{n-1}(t_{n-1})] \quad (2 \leq n \leq \lfloor \frac{T}{T_{min}} \rfloor) \end{aligned}$$

この漸化式の結果から分析区間における最小累積区間距離は $\min_n(D_n(T))$ 、最適分割数 p は $\arg \min_n(D_n(T))$ で求めることができる。そして、分析区間 T における最適時間セグメントの i 回目の分割時刻 τ_i は以下のように求めることができる。

$$\begin{aligned} \tau_p &= T \\ \tau_{p-1} &= \arg \min_t [D(\tau_p - t, \tau_p) + D_{p-1}(t)] \\ \tau_{p-2} &= \arg \min_t [D(\tau_{p-1} - t, \tau_{p-1}) + D_{p-2}(t)] \\ &\vdots \\ \tau_2 &= \arg \min_t [D(\tau_3 - t, \tau_3) + D_2(t)] \\ \tau_1 &= \arg \min_t [D(\tau_2 - t, \tau_2) + D_1(t)] \end{aligned}$$

これらの計算により最適時間セグメントを抽出することが可能となる。

第4章 識別実験

4.1 実験概要

提案した楽器・音高識別処理手法、最適時間セグメント抽出処理手法の精度を確かめるために、様々なパターンの楽音（テストデータ）を入力して演奏情報の抽出を行い、入力楽音の演奏情報との比較を行う。このとき、第2章で提案した3つの楽器・音高識別手法それぞれについて演奏情報の抽出を行い、各手法の精度の比較も行う。

テストデータは以下の実験 1-1 ~ 2-2 のように単音でオーバーラップの無い楽音、混合音でオーバーラップの無い楽音、単音でオーバーラップの有る楽音、混合音でオーバーラップの有る楽音の4種類を用意した。

実験 1-1 単音 オーバーラップ無し

実験 1-2 混合音 オーバーラップ無し

実験 2-1 単音 オーバーラップ有り

実験 2-2 混合音 オーバーラップ有り

4.2 オーバーラップ

ここでは、オーバーラップを図 4.1 に示すように、ある音高の音が鳴り始めてから鳴り終る間に、別の音高の音が鳴り始める状態と定義する。また、図の見方であるが、横軸が時間、縦軸が音高、色の濃淡がゲイン値を示しており、図 4.1 では 0.0 秒からおよそ 0.9 秒までピアノの $E4$ の音が、およそ 0.4 秒からおよそ 1.2 秒までピアノの $A4$ の音が鳴っているのを表している。

一般的な楽音においてオーバーラップは頻繁に出現する。しかしながら、本研究で提案する手法ではオーバーラップに対応していないので今後の検討が必要となる。今回は提案手法においてオーバーラップの影響がどの程度あるのかを調べるために実験 2-1, 2-2 を行った。

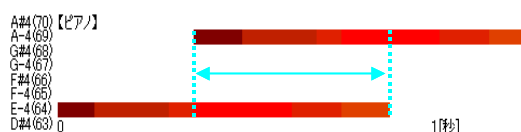


図 4.1: オーバーラップの例

4.3 テストデータ

実験 1-1 では MIDI 楽音作成ソフトを使ってピアノ和音 (C4,E4,G4) の数秒後に別のピアノ和音 (D4,F4,A4) が鳴る MIDI データを作成し演奏データとした (図 4.2 下段参照)。実験 1-2 では実験 1-1 で作成した演奏データにフルート (E4,F4) を加えて混合音とした (図 4.2 上下段参照)。

また、RWC 研究用データベースに収録してある編成バリエーション楽曲には、その楽曲を MIDI で再現したデータも収録されている。そこで、ピアノとフルートで演奏されている MIDI データから、ピアノ部分を抽出した MIDI データを実験 2-1 用 (図 4.3 下段参照) の演奏データとし、ピアノとフルートで演奏されているものを実験 2-2 用 (図 4.3 上下段参照) の演奏データとした。

以下の図 4.2,4.3 は実験で用いた演奏データの楽器 (上段:フルート, 下段:ピアノ)、音高、ゲイン値を図に表したものである。図の縦軸は音高を、横軸は時間、色の明るさがゲイン値を表している。

ここで、テストデータに用いる音源とテンプレートに用いる音源を同じにするために、MIDI データをそのままテストデータにするのではなく、MIDI データから楽器、音高、ゲイン値、鳴り始めの時間、鳴り終りの時間の 5 つの演奏情報を抽出し、音源をテンプレート作成時に用いた単音データと差し替えることでテストデータとした。

このようにすることで、音色が正しいときのマッチング手法の精度を確かめることができる。

4.4 楽器データ

テンプレート、テストデータの作成には以下の単音データと楽音データを用いる。

4.4.1 単音データ（混合音モデル作成用）

RWC 研究用データベースに収録されている楽器音データベースの中から 4.4.2 の楽音データの曲中に使われる楽器を全て選び、選んだ全楽器の全音階毎の単音を単音データとする。そして、これら全ての単音データを基に混合音モデル（テンプレート）を生成する。

4.4.2 楽音データ

楽音データには RWC 研究用データベース [4] に収録されているジャズ音楽データベースの中から編成バリエーション楽曲 35 曲の MIDI を用いる。

この編成バリエーション楽曲は、同一曲が異なる編成で演奏される事例を得ることを目的として収録されたもので、編成バリエーションを変えることでタスクの難易度を調節できる。

ここでは簡単な分析データとして“ ピアノソロ ”, ピアノとフルートの“ デュオ ”の編成バリエーションを用いる。

4.4.3 編成バリエーション楽曲

編成バリエーション楽曲とは、同一曲が異なる編成で演奏される事例を得ることを目的として収録されたもので、スタンダード風の楽曲を 5 曲新規作曲（メロディーとコード進行を決定）し、それぞれを以下の 7 編成でモダンジャズ風に演奏したものである。

1. ピアノソロ
2. ギターソロ
3. デュオ（ピアノ + 1 楽器）
4. ピアノトリオ（ピアノ、ベース、ドラム）
5. 上記ピアノトリオ + テナーサックスまたはトランペット

6. オクテット (上記ピアノトリオ + ギター + アルトサックス + バリトンサックス + テナーサックス × 2)
7. 上記ピアノトリオ + ビブラフォンまたはフルート

そして、編成バリエーション楽曲 35 曲それぞれに対し MIDI 符号化されたデータが存在する。

4.5 実験条件

本実験では 4.4.1、4.4.2 で述べた楽器データを標本周波数 16000Hz にダウンサンプリングして使用する。量子化ビット数は 16bits である。波形レベルで高域強調した後、窓長 0.25 秒、分析周期 0.05 秒、分析窓に hamming 窓を使用して FFT によりパワースペクトルを抽出した。また、使用するスペクトルは 2000Hz までと帯域制限をしている。

各実験において楽器・音高識別用の混合音モデル (テンプレート) はテストデータで使用する楽器・音高を全て網羅できるようにピアノの $D2 \sim A\sharp 5$ 、フルートの $D4 \sim A\sharp 4$ の単音データを用いて作成している。

最適時間セグメント抽出処理においての可変長の時間セグメントの最短時間長を 0.2 秒、最長時間長を 2 秒、シフト時間長を 0.05 秒とした。

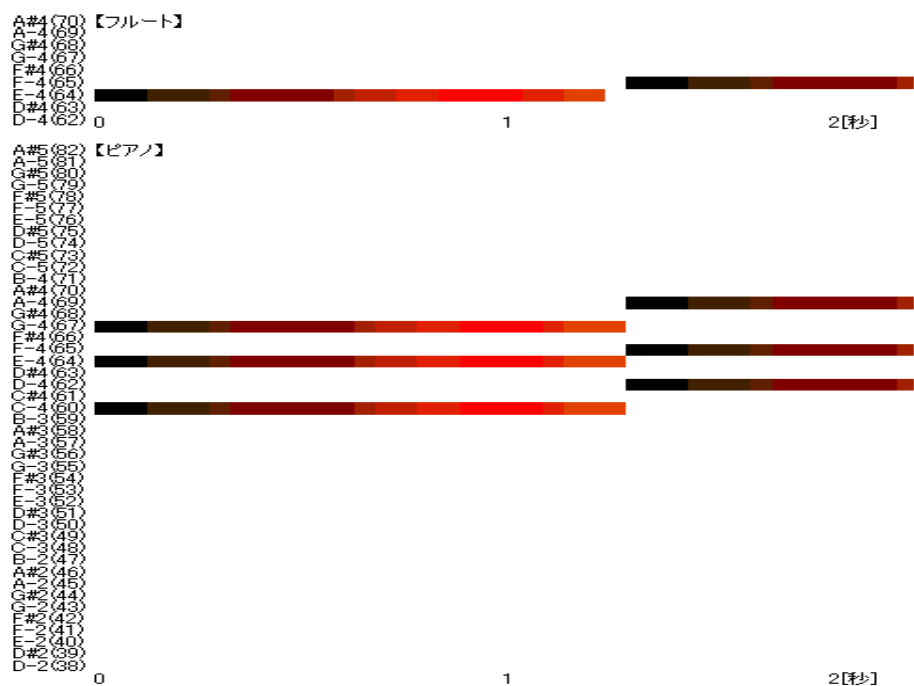


図 4.2: 実験 1-1、1-2 用テストデータ

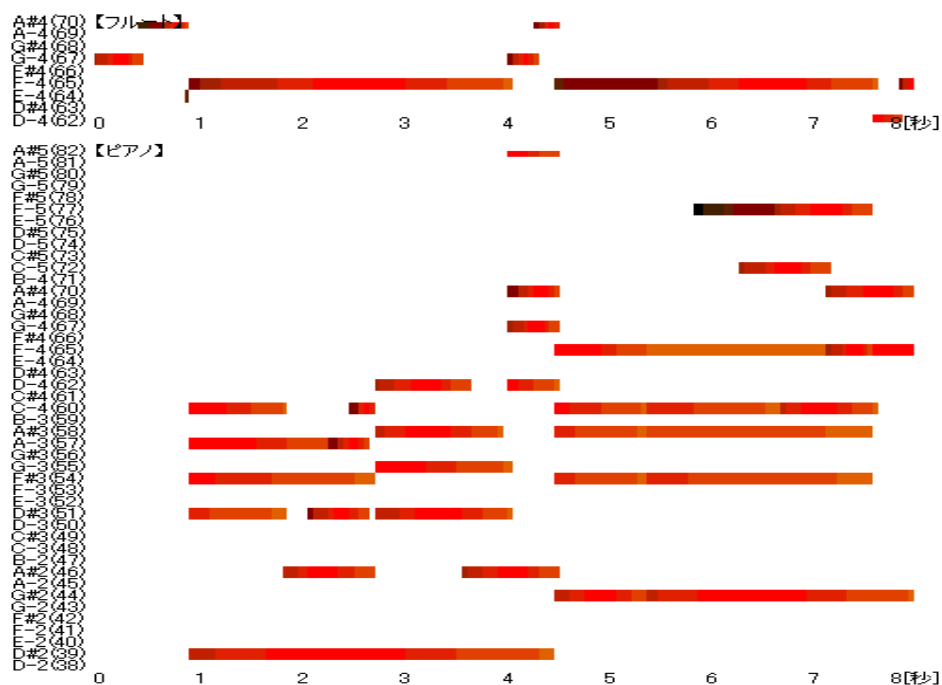


図 4.3: 実験 2-1、2-2 用テストデータ

第5章 実験結果

実験 1-1 ~ 2-2 のテストデータに対し、手法 1 ~ 3 による識別実験を行った。各テストデータ、各手法による結果を以下 5.1 ~ 5.4 に示す。

5.1 実験 1-1 結果

図 5.1 ~ 5.3 に実験 1-1 の結果を示す。縦の線は最適時間セグメント抽出処理によって抽出された最適時間セグメントの分割時刻である。

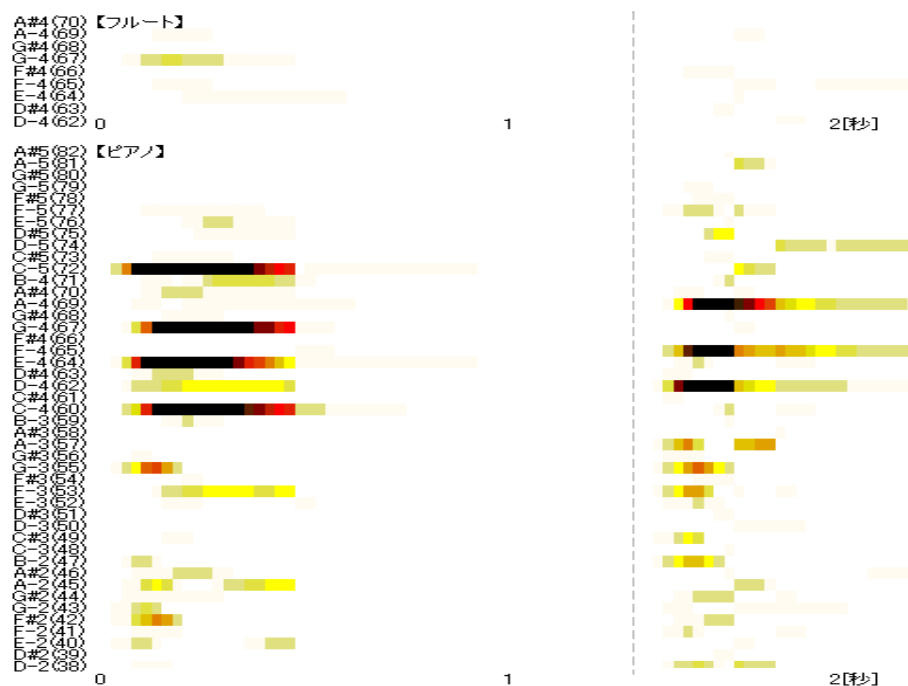


図 5.1: 実験 1-1 手法 1

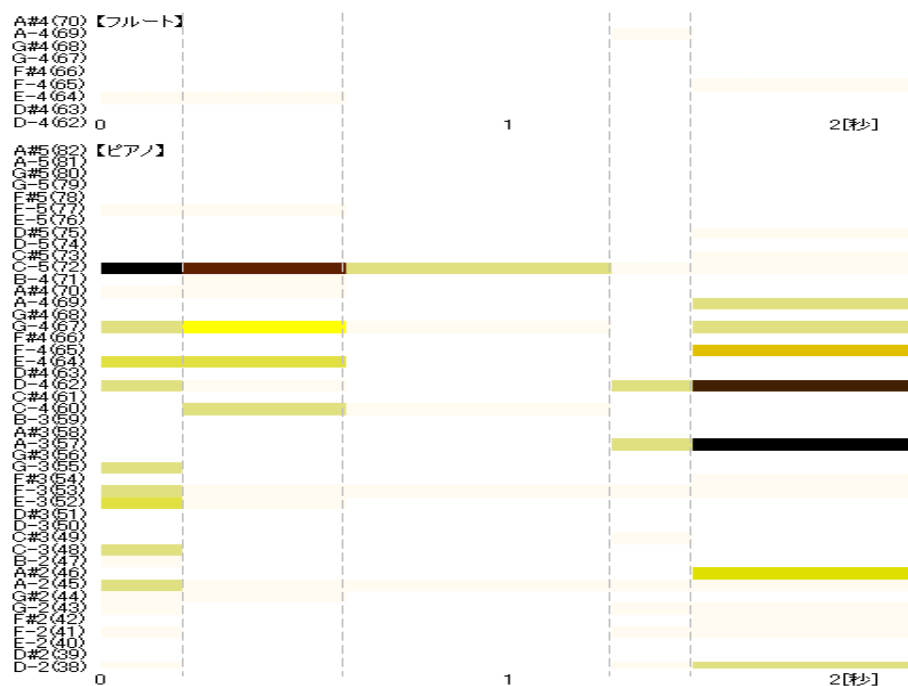


図 5.2: 実験 1-1 手法 2

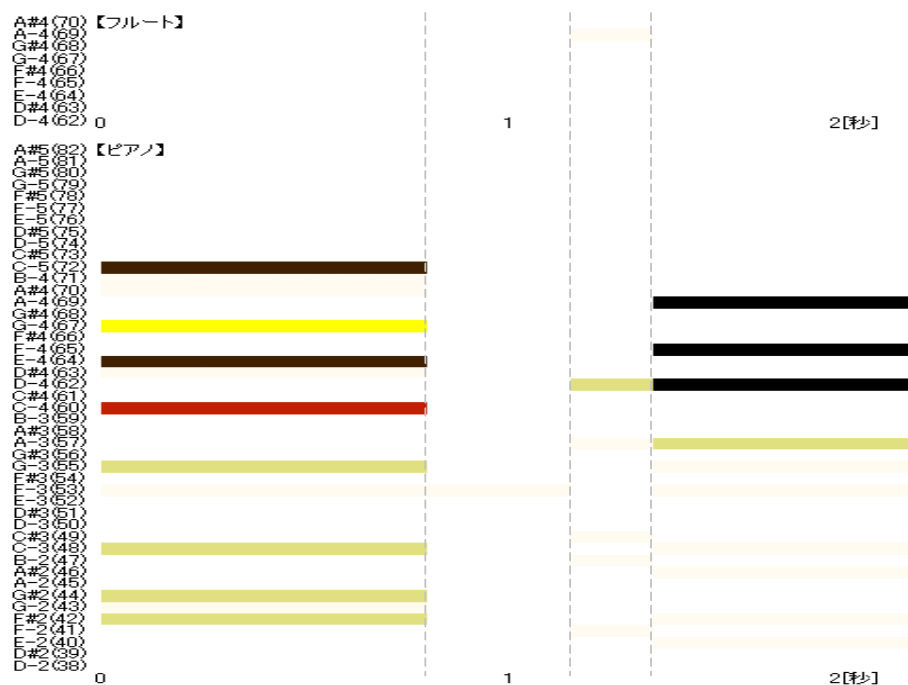


図 5.3: 実験 1-1 手法 3

5.2 実験1-2 結果

図5.4～5.6に実験1-2の結果を示す。

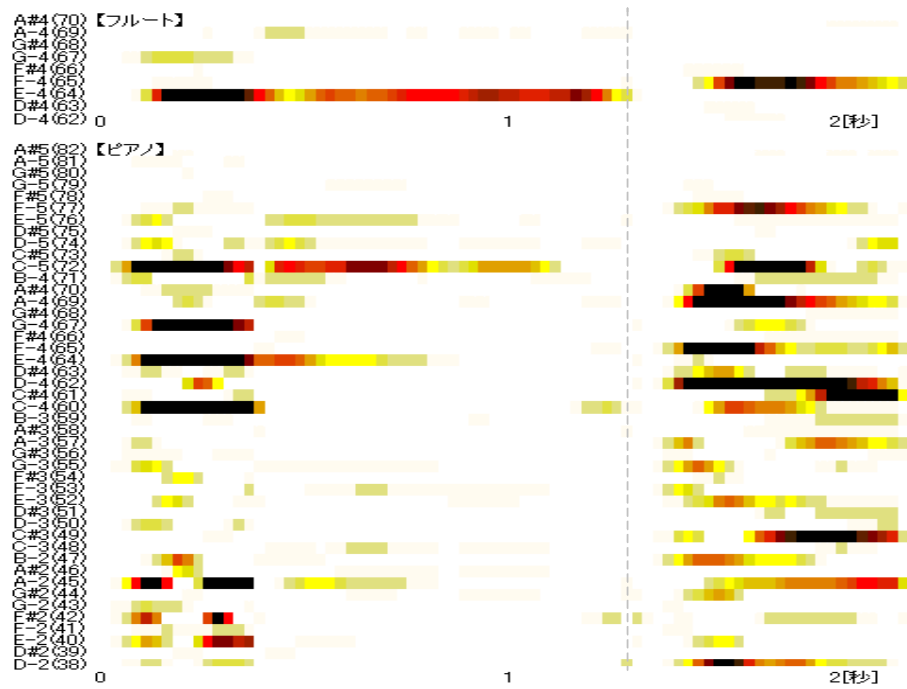


図 5.4: 実験 1-2 手法 1

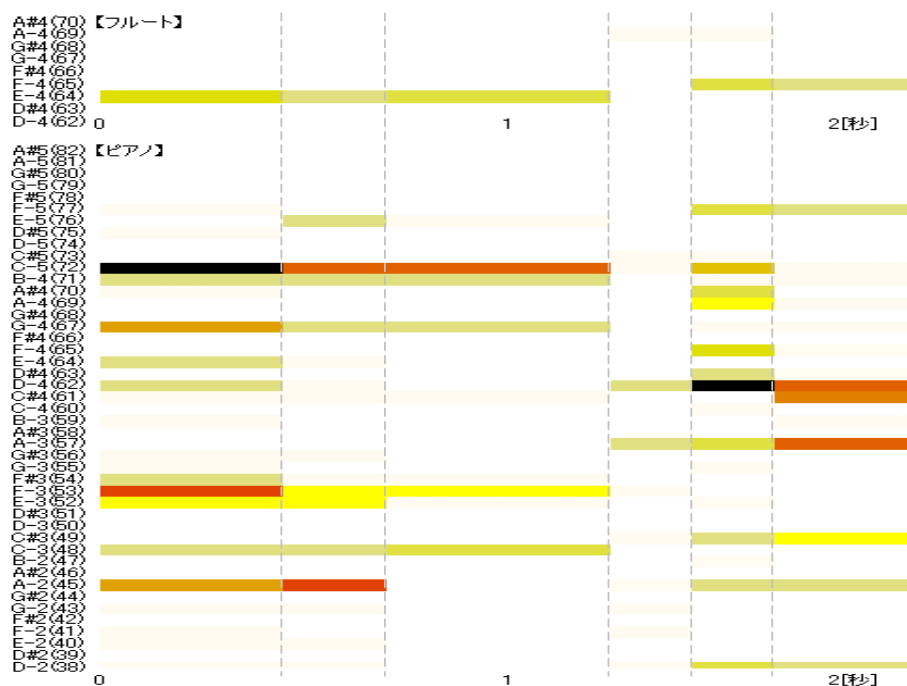


図 5.5: 実験 1-2 手法 2

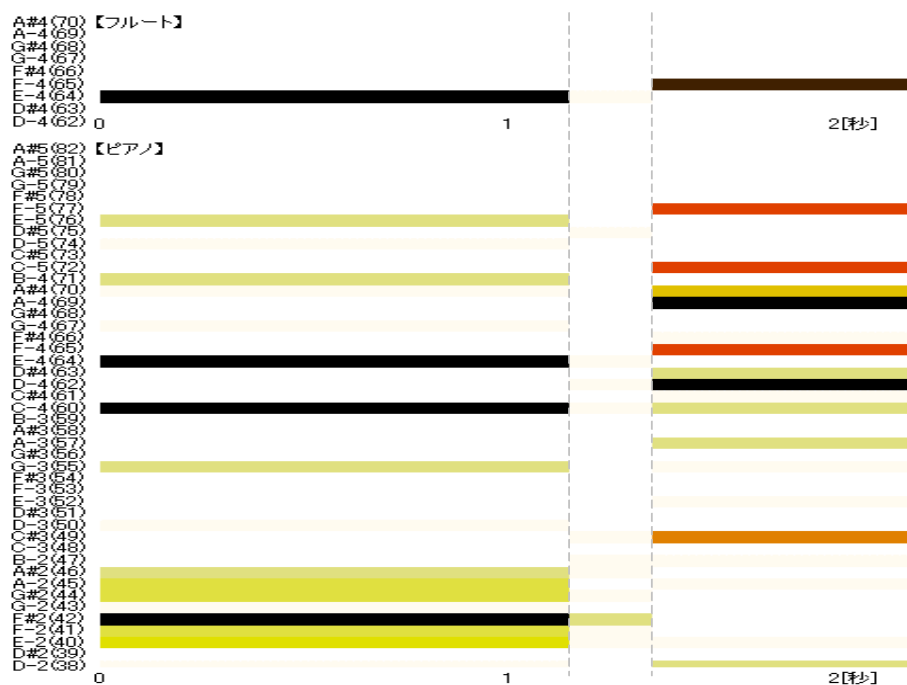


図 5.6: 実験 1-2 手法 3

5.3 実験 2-1 結果

図 5.7～5.9 に実験 2-1 の結果を示す。

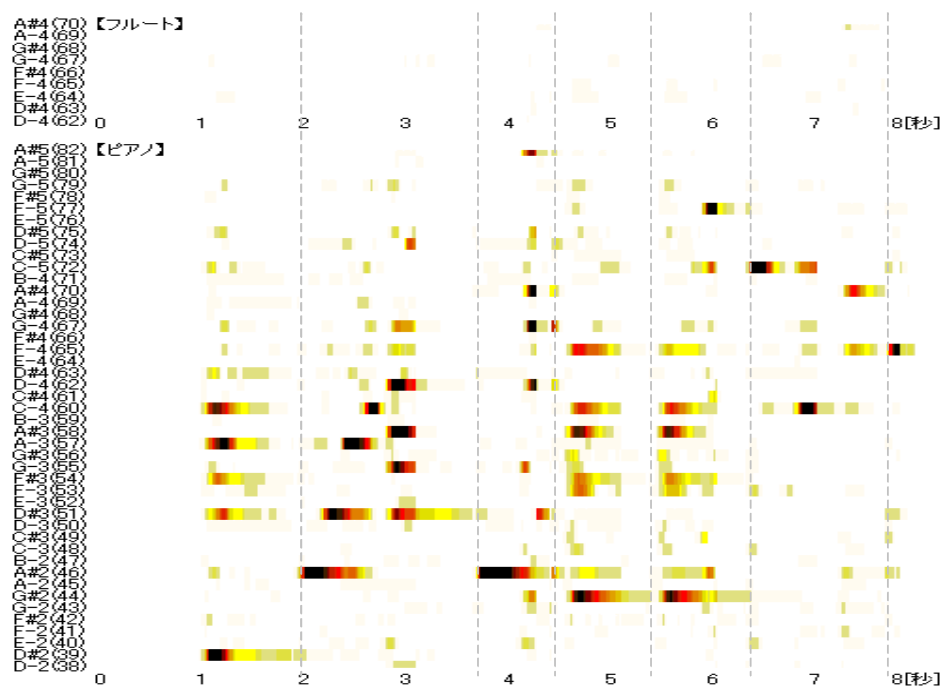


図 5.7: 実験 2-1 手法 1

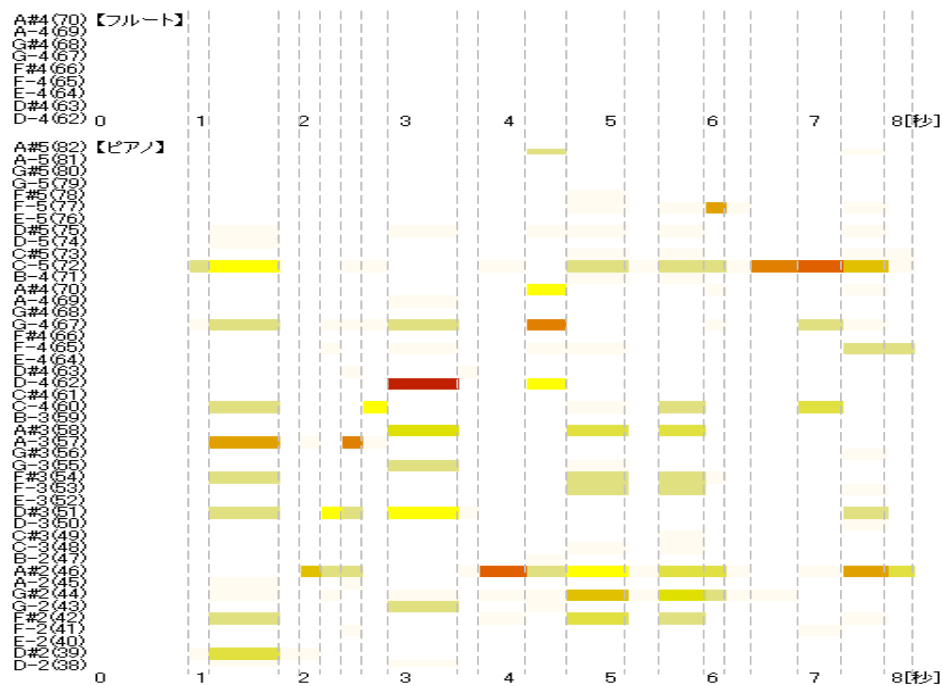


図 5.8: 実験 2-1 手法 2

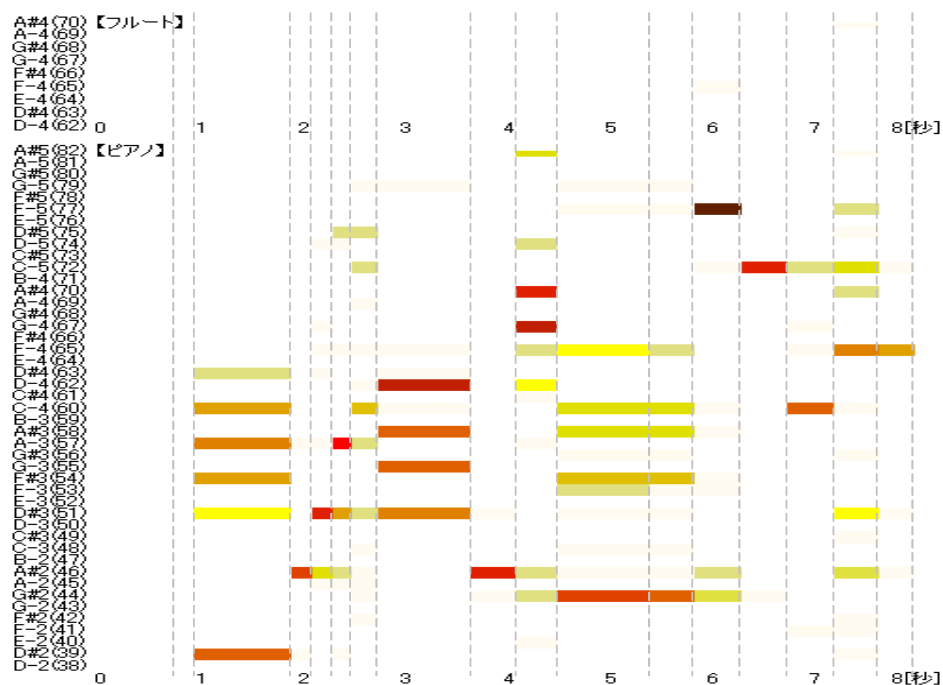


図 5.9: 実験 2-1 手法 3

5.4 実験2-2 結果

図 5.10 ~ 5.12 に実験 2-2 の結果を示す。

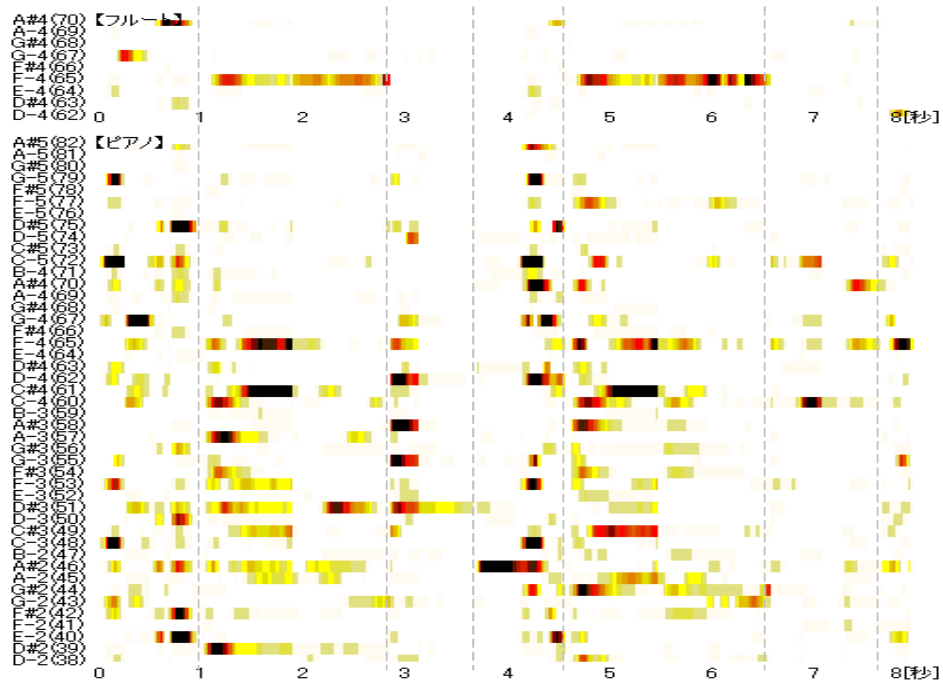


図 5.10: 実験 2-2 手法 1

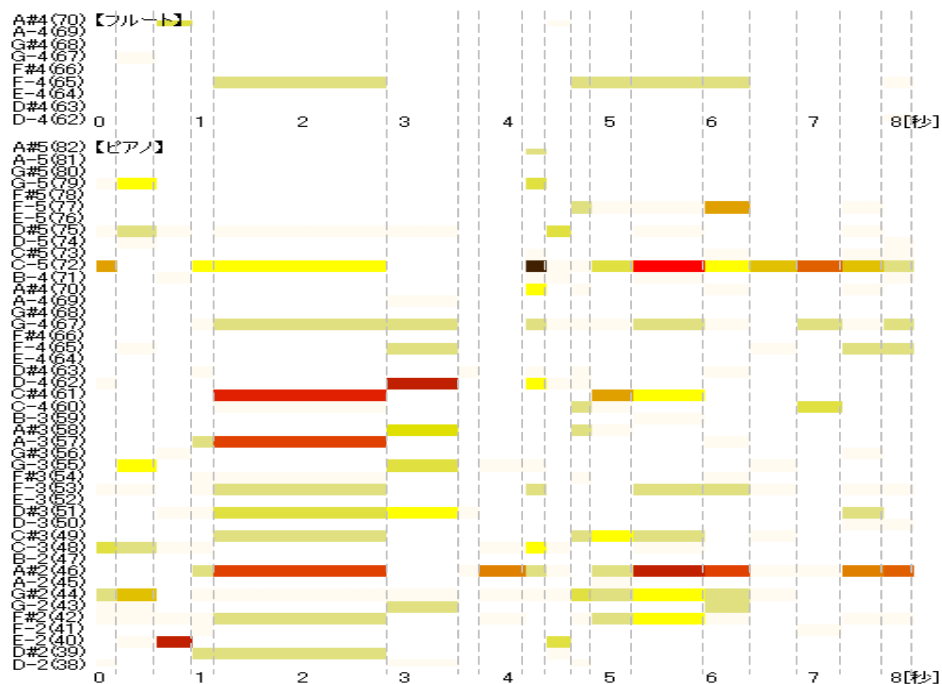


図 5.11: 実験 2-2 手法 2

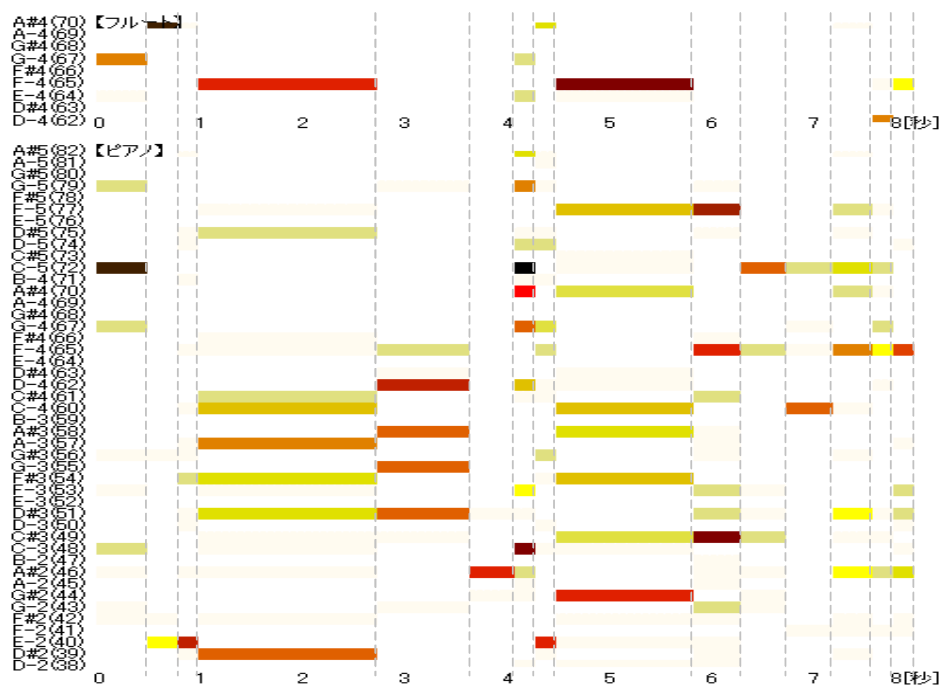


図 5.12: 実験 2-2 手法 3

5.5 平均スペクトル距離

入力スペクトル s と距離が最小となったときの混合音モデル \hat{s} との平均スペクトル距離を各実験、各手法について表 5.1 にまとめた。時刻 i における入力スペクトルの周波数 ω における値を $s_i(\omega)$ 、時刻 i の入力スペクトルとの距離を最小としたときの混合音モデルのスペクトルの周波数 ω における値を $\hat{s}_i(\omega)$ 、スペクトル時系列長を N とし、平均スペクトル距離 D_{ave} を以下の式 (5.1) で算出した。

$$D_{ave} = \frac{1}{N} \sum_i^N \sqrt{\frac{1}{n} \sum_{\omega}^n (s_i(\omega) - \hat{s}_i(\omega))^2} \quad (5.1)$$

表 5.1: 平均スペクトル距離

| | 実験 | | | |
|------|-------|--------|-------|-------|
| | 1-1 | 1-2 | 2-1 | 2-2 |
| 手法 1 | 5.83 | 13.54 | 2.45 | 7.63 |
| 手法 2 | 34.62 | 101.75 | 20.14 | 41.22 |
| 手法 3 | 3.49 | 10.45 | 1.98 | 5.91 |

第6章 考察

実験 1-1 の結果について多少の誤差はあるも、図 4.2 の約 1.4 秒のところで次の和音 ($D4, F4, A4$) が立ち上がっているように、各手法共 (図 5.1, 5.2, 5.3 参照) 1.4 秒付近で時間セグメントの分割が行われていることが確認できる。ここで、手法 2・3 について時間セグメントがそれぞれ 5 セグメント、4 セグメントに分割されているが、これは手法 2・3 が最適時間セグメント抽出処理において、分割数を増やすほど累積区間距離が最小となりやすいためである。この 2 つの手法について分割数 2 を与えて処理を行ったところ、約 1.4 秒付近で時間セグメントの分割が行われていることが確認できている。

楽器・音高識別処理では手法 1・3 は誤った音高も抽出しているものの、正解の音高を漏れなく抽出している。また、誤って抽出した音高は正解の音高とオクターブの関係になっていることが確認できる。これはオクターブ関係の 2 音には、楽器の倍音構造により、同じ周波数に大きいパワーを持ちやすいためであると考えられる。

実験 1-1 のテストデータを各手法を用いて演奏情報を抽出し、MIDI データを作成し聴いてみた結果、どの手法共元の楽音に近い楽音を得ることができた。

実験 1-2 は実験 1-1 にフルートを加えたテストデータ (図 4.2) となっているが、各手法共正しいフルートの音高を抽出していることが図 5.4, 5.5, 5.6 から確認できる。但し、ピアノの抽出が実験 1-1 のときと比べ乱れているのも同時に確認できる。これは、フルートに含まれる周波数成分をピアノの各音高のテンプレートで補おうとした結果だと考えることができる。

実験 1-2 のテストデータを各手法を用いて演奏情報を抽出し、MIDI データを作成し聴いてみた結果、どの手法共元の楽音に近い楽音を得ることができたが、実験 1-1 のピアノ単音のときの結果と比べると、ピアノの演奏が乱れているのが確認できた。

実験 2-1 は実際の楽曲のピアノ部分のテストデータであり、複雑な和音やオーバーラップを含んでいるが、各手法共ある程度正しい時間に正しい音高が抽出されているのを確認できる (図 4.3, 5.7, 5.8, 5.9 参照)。

実験 2-1 のテストデータを各手法を用いて演奏情報を抽出し、MIDI データを作成し聴いてみた結果、どの手法共元の楽音に近い楽音を得ることができた。特に手法 1・3 において

は表 5.1 からスペクトル距離が元の楽音に近いことが分かるように、かなり元の楽音に近い楽音を得ることができた。

実験 2-2 は実験 2-1 にフルートの演奏を加えたものであるが、これは実験 1-1 と実験 1-2 の関係と同様、フルートの抽出はできるが、ピアノの抽出が乱れるということが図 4.3, 5.7, 5.8, 5.9 と図 4.3, 5.10, 5.11, 5.12 を比較することで確認することができる。これも実験 1-2 と同様、フルートに含まれる周波数成分をピアノの各音高のテンプレートで補おうとした結果だと考えることができる。

表 5.1 から各手法について比較すると、手法 3 のスペクトル距離が最小であることが分かる。これは手法 3 のテンプレートに用いたスペクトルが手法 1・2 と比べ多いためと考えられる。また、実験 1-1、2-1 と実験 1-2、2-2 のスペクトル距離から、混合音では単音と比べ識別時のマッチング誤差が大きく増加することが確認できた。一般的に入力信号を混合音にする場合、どの従来研究からも精度が落ちることから、本手法でも精度が落ちたのだと考えられる。

今回提案した手法では、4.2 で述べたオーバーラップの問題を考慮しなかったが、図 5.7 の約 1.9 秒から約 2.6 秒にかけて、オーバーラップに似た出力が得られた。これは手法 1 のゲイン値が時間セグメント内で可変なため出力されたと考えられる。また、図 5.9 の約 1.9 秒から約 2.6 秒にかけて、時間セグメントを細かく区切ることでオーバーラップのような出力が見られた。これらを考慮すれば今後オーバーラップの問題を解決できるのではないかと考えている。

第7章 まとめ

今回の識別実験の結果から手法1・3において入力楽音と全く同じ演奏情報は抽出されなかったが、ある程度正解の楽器・音高を抽出できることが確認できた。実際に出力した演奏情報からゲイン値を補正してMIDIデータを作成し聴いてみたところ、元の楽音にある程度似た楽音を得られたことから、本手法は楽音の楽器・音高識別において有効な手法であると考えることができる。

参考文献

- [1] 李 而立: “ 心理的距離尺度に基づく実音色から有限音色空間への写像 ”, 早稲田大学 2003 年度卒業論文, 2004.
- [2] 今井 義徳: “ クラシック音楽を対象とした一般調和解析による採譜支援法の提案 ”, 早稲田大学 2003 年度修士論文, 2004.
- [3] 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一.“ RWC 研究用音楽データベース: 音楽ジャンルデータベースと楽器音データベース ”, 情報処理学会 音楽情報科学研究会 研究報告 2002-MUS-45-4, Vol.2002, No.40, pp.19-26, May 2002.
- [4] 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一.“ RWC 研究用音楽データベース: クラシック音楽データベースとジャズ音楽データベース ”, 情報処理学会 音楽情報科学研究会 研究報告 2002-MUS-44-5, Vol.2002, No.14, pp.25-32, Feb 2002.
- [5] 木下 智義: “ 音楽音響信号を対象とした自動採譜処理に関する研究 ”, 博士論文, 東京大学大学院, 工学系研究科, 2000.
- [6] Kunio Kashino and Hiroshi Murase: “ A music stream segregation system based on adaptive multiagents ”, In Proceedings of the 15th International Joint Conference on Artificial Intelligence, Vol.2, pp.11126-1131, Aug. 1997.
- [7] David Rosenthal: “ Machine Rhythm: Computer Emulations of Human Rhythm Perception ”, PhD thesis, Massachusetts Institute of Thechnology, 1992.
- [8] 後藤 真孝: “ 音楽音響信号を対象としたリアルタイムビートトラッキングに関する研究 ”, PhD thesis, 早稲田大学大学院理工学研究科, Mar. 1998.
- [9] 北原 鉄朗, 後藤 真孝, 奥乃 博: “ 楽器音オントロジー作成のための楽器音特徴抽出 ”, 情報処理学会 第 62 回全国大会, 4M-5, Mar. 2001.

- [10] 北原 鉄朗, 後藤 真孝, 奥乃 博: “ 音高による音色変化に着目した音源同定手法 ”, 情報処理学会 音楽情報科学研究会 研究報告, 2001-MUS-40-2, Vol.2001, No.45, pp.7-14, May 2001.
- [11] 北原 鉄朗, 後藤 真孝, 奥乃 博: “ 楽器音を対象とした音源同定: 音高による音色変化を考慮する識別手法の検討 ”, 情報処理学会 音楽情報科学研究会 研究報告, 2002-MUS-46-1, Vol.2002, No.63, pp.1-8, July 2002.
- [12] 北原 鉄朗, 後藤 真孝, 奥乃 博: ” 音色空間の音高依存性を考慮した楽器音の音源同定 ”, 日本音響学会 2002 年秋季研究発表会 講演論文集, 1-1-4, pp.643-644, Sept. 2002.
- [13] 北原 鉄朗, 後藤 真孝, 奥乃 博: ” 音響的類似性に基づく楽器音の階層的クラスタリング ”, 情報処理学会 第 65 回全国大会, 1P-1, Mar. 2003.
- [14] 北原 鉄朗, 後藤 真孝, 奥乃 博: ” 音響的特徴に基づく楽器の階層表現の獲得とそれに基づくカテゴリーレベルの楽器音認識の検討 ”, 情報処理学会 音楽情報科学研究会 研究報告, 2003-MUS-51-9, Vol.2003, No.82, pp.51-58, Aug. 2003.
- [15] 片寄 晴弘, 井口 征士: “ 知的採譜システム ”, 人工知能学会誌, Vol.5, No.1, pp.59-66, 1990.
- [16] 三輪 多恵子, 田所 嘉昭, 斎藤 努: “ くし形フィルタを利用した採譜のための異楽器音中のピッチ推定 ”, 電子情報通信学会論文誌, Vol.J81-DII, No.9, pp.1965-1974, Aug. 1998.
- [17] 柏野 邦夫, 中臺 一博, 木下 智義, 田中 英彦: “ 音楽情景分析の処理モデル OPTIMA における単音の認識 ”, 電子情報通信学会論文誌, Vol.J79-DII, No.11, pp.1751-1761, Nov. 1996.
- [18] 柏野 邦夫, 村瀬 洋: “ 適応型混合テンプレートを用いた音源同定 ”, 電子情報通信学会論文誌, D-II Vol.J81-D-II No.7 pp.1510-1517 July 1998.
- [19] 日比啓文, 半田伊吹, 武藤誠, 坂井修一, 田中英彦. “ 音色空間を用いた音源分離システム ”, 情報処理学会 第 62 回全国大会, 慶應義塾大学矢上キャンパス, No.5E-4, Mar. 2001
- [20] 安藤 由典: “ 楽器の音響学 ”, 音楽之友社, 1996.
- [21] R.A.Rasch and R. Plomp (宮坂 栄一訳): “ 音楽の知覚 ”, 音楽の心理学 (上) 第 1 章, 西村書店, 1987.

- [22] J.C.Risset and D.L.Wessel (宮坂 栄一訳): “ 分析と合成による音色の探求 ”, 音楽の心理学 (上) 第2章, 西村書店, 1987.
- [23] 境 久雄, 中山 剛: “ 聴覚と音響心理 ”, コロナ社, 1978.

謝辞

本研究を進めるにあたり、研究環境を整え、ご指導をいただいた白井克彦先生には心より感謝致します。

音響信号班のチームゼミにおいて基礎技術指導や的確なアドバイスをして下さった誉田雅彰先生、樽松明先生、大川茂樹先生、金子格先生、出口幸子先生には大変感謝しております。

ご自分の研究で忙しい中、指導をしてくれたり、初歩的な質問にも丁寧に答えて下さった博士過程2年の谷口徹さんには大変感謝しており、その念に尽きません。

修士論文という大変な作業を共に頑張ってきた白井研究室のM2の皆に感謝致します。特に、苦労を共にした、音声班M2の大久保君、山本君、小林さんには感謝の気持ちでいっぱいです。

また、就職活動の中、夏合宿、合同発表会の準備などのお世話をしてくれたM1の皆、一緒に論文を書いたB4の皆には感謝します。

発表や資料提出の直前など、研究室に残り徹夜の作業をし、互いに励まし合い、共に苦労を重ねた同輩のみんなとの辛くもあり楽しくもあった日々を生涯忘れません。

最後に、本大学院への進学に理解を示し、6年間もの学業生活を支え、温かく見守って下さった両親、弟、妹、親族の方に深く感謝致します。

2005年2月

椿 雅也